



VERIFICATION OF TRANSLATION

Patent Application No. 304059/96

I, Hidejiro TANIGAWA of 9-3-1906, Hikarigaoka
3-chome, Nerima-ku, Tokyo 179, Japan, am the
translator of the documents attached and I state that
the following is a true translation to the best of my
knowledge and belief of Japanese Patent Application
No. 304059/96.

DATED this 21st day of October, 1999

Signature of translator

Hidejiro Tanigawa

Hidejiro TANIGAWA

[TITLE OF DOCUMENT] REQUEST FOR PATENT

[DOCKET NUMBER] 96475

[DATE SUBMITTED] October 30, 1996

[TO] Director General of Patent Office

[INTERNATIONAL PATENT CLASSIFICATION] C12N 15/00

[TITLE OF THE INVENTION] cDNA Fragment of Causative Gene of Spinocerebellar Ataxia Type 2

[NUMBER OF CLAIMS] 2

[INVENTOR]

[ADDRESS OR DOMICILE] 1-10-45, Kobari, Niigata-shi, Niigata, JAPAN

[NAME] Shoji TSUJI

[INVENTOR]

[ADDRESS OR DOMICILE] A-203 Advantage, 98-1, Oshima, Niigata-shi, Niigata, JAPAN

[NAME] Kazuhiro SANPEI

[APPLICANT FOR PATENT]

[IDENTIFICATION NUMBER] 390037006

[NAME] SRL, INC.

[REPRESENTATIVE] Toshiyuki KONDO

[PHONE] (03)3344-6511



[ATTORNEY]

[IDENTIFICATION NUMBER] 100088546

[PATENT ATTORNEY]

[NAME] Hidejiro TANIGAWA

[PHONE] 03-3238-9182

[INDICATION OF FEE]

[AMOUNT] 21000

[LIST OF DOCUMENTS SUBMITTED]

[TITLE OF DOCUMENT] Specification 1

[TITLE OF DOCUMENT] Abstract 1

[TITLE OF DOCUMENT] Drawing 1

[GENERAL POWER OF ATTORNEY NO.] 9205311



[TITLE OF DOCUMENT] SPECIFICATION

[TITLE OF THE INVENTION] cDNA Fragment of Causative Gene of
Spinocerebellar Ataxia Type 2

[CLAIMS]

5 [Claim 1] A DNA fragment comprising a DNA region encoding an
amino acid sequence shown in SEQ ID NO: 1 (provided that the number of repeat
units of Gln from the 166th to 188th amino acid varies between 15 and 100).

[Claim 2] The DNA fragment according to claim 1, wherein said DNA
region is the region from 49nt to 3987nt (provided that the number of repeat units of
10 CAG or CAA in the region from the 543nt to 612nt varies between 15 and 100, and
that the CAA in this region may be CAG).

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Technical Field of the Invention]

15 The present invention relates to cDNA fragments of the causative gene of
spinocerebellar ataxia type 2 (hereinafter also referred to as "SCA2").

[0002]

[Prior Art]

20 SCA2 is an autosomal dominant, neurodegenerative disorder that affects the
cerebellum and other areas of the central nervous system.

[0003]

It has recently been discovered that the causative genes of
5 neurodegenerative diseases including dentatorubral-pallidoluysian atrophy
(DRPLA) have more CAG repeats than the normal genes. That is, the numbers of
25 CAG repeats in the causative genes of the neurodegenerative diseases are 37 to 100,
while those in the normal genes are less than 35.

[0004]

It has been suggested that the causative gene of SCA2 has an increased number of CAG repeats (Trottier, Y. et al. Nature, 378, 403-406 (1995)). However, since the causative gene of SCA2 has not been identified, and since its nucleotide sequence has not been determined, SCA2 cannot be diagnosed by a genetic assay.

5 [0005]

[Problems Which the Invention Tries to Solve]

An object of the present invention is to provide a sequence-determined cDNA fragment of the causative gene of SCA2.

[0006]

10 [Means for Solving the Problems]

The present inventors intensively studied to discover a *Tsp* E1 fragment with a size of 2.5 kb in which the number of CAG triplet is increased only in SCA2 patients, and partial sequence thereof was determined. Human cDNA library was screened using as probes the oligonucleotides that respectively hybridize with the regions between which the CAG triplet repeats are interposed, and a cDNA fragment which hybridizes with both of these two probes was cloned. Using this cDNA fragment as a probe, human cDNA library was screened and a plurality of cDNA fragments which hybridize with the probe were cloned. Sequencing the cDNA fragments revealed that these cDNA fragments overlap with each other. To
15
20 sequence the 5'-end and 3'-end regions, RACE (rapid amplification of cDNA ends) was performed. Further, to sequence the 5'-end region, RT-PCR was performed, thereby succeeding in sequencing the full length of the cDNA of the causative gene of SCA2.

[0007]

25 That is, the present invention provides a DNA fragment comprising a DNA region encoding an amino acid sequence shown in SEQ ID NO: 1 (provided that the number of repeat units of Gln from the 166th to 188th amino acid varies between 15

and 100).

[0008]

[Modes for Carrying Out the Invention]

As described above, the DNA fragment according to the present invention
5 comprises the DNA region encoding the amino acid sequence shown in SEQ ID NO:
1 in the SEQUENCE LISTING, provided that the number of repeating units of Gln
from the 166th to the 188th amino acid varies between 15 and 100. The number of
this repeat units is 15 to 25 in normal individuals and 35 to 100 in SCA2 patients.
As is well-known, due to degeneration, there are a plurality of codons encoding one
10 amino acid, and any DNAs which encode the amino acid sequence shown in SEQ ID
NO: 1 are included within the scope of the present invention. The nucleotide
sequence actually determined in the Examples below is shown in SEQ ID NO:1 and
in Figs. 1-4. The way how the nucleotide sequence was determined and the fact
that the cDNA having this nucleotide sequence is the cDNA of the causative gene of
15 SCA2 are detailed in the Examples below.

[0009]

Since the nucleotide sequence of the nucleic acid fragment according to the
present invention was determined by the present invention, the nucleic acid fragment
may be cloned by utilizing amplification by PCR using human cDNA library as a
20 template. In cases where it is difficult to amplify the nucleic acid fragment by a
single PCR, the nucleic acid fragment may be divided into a plurality of regions and
the PCR products may be ligated by a conventional method so as to clone the nucleic
acid fragment.

[0010]

25 [Example]

The present invention will now be described more concretely by way of
examples thereof. It should be noted that the present invention is not restricted to

the following examples.

[0011]

Example 1 Preparation of (CAG)₅₅ Probe

A genomic DNA segment of DRPLA gene containing a CAG repeat with 55
 5 repeat units was amplified from the genomic DNA of a patient with DRPLA
 (Koide, R. et al., Nature Genet., 6, 9-13 (1994)) and was subcloned into a plasmid
 vector, pT7Blue T(*p-2093*). The *p-2093* plasmid contains the (CAG)₅₅ and the
 flanking sequences. That is, the plasmid contains the sequence of 5'-CAC CAC
 CAG CAA CAG CAA (CAG)₅₅ CAT CAC GGA AAC TCT GGG CC-3'. Using a
 10 pair of oligonucleotides 5'-CAC CAC CAG CAA CAG CAA CA-3' and 5'-biotin-
 GGC CCA GAG TTT CCG TGA TG-3', PCR was performed in a total volume of 16
 µl containing 10 mM Tris-HCl, pH8.3, 50 mM KCl, 1.5 mM MgCl₂, 2M N,N,N-
 trimethylglycine, 0.1 mM TTP, 0.1 mM dCTP, 0.1 mM dGTP, 9.25 MBq of [α -
³²P]dATP (222 TBq/mmol), 0.5 µM each of the two primers, 0.3 ng of plasmid DNA
 15 (*p-2093*) and 2.0 U of Taq DNA polymerase (Takara Shuzo, Kyoto, Japan). After
 an initial 2-min. denaturation at 94°C, PCR was performed for 30 cycles consisting
 of 1-min. denaturation at 94°C, 1-min. annealing at 54°C and 3-min. extension at
 72°C, followed by a final extension at 72°C for 10 min.

[0012]

20 A single-stranded (CAG)₅₅ probe was isolated using streptavidin-coated
 magnetic beads (Dynabeads M-280, Streptavidin; Dynal AS, Oslo, Norway) on which
 20 µl of streptavidin is coated. That is, after washing of the PCR products
 immobilized on the magnetic beads with 40 µl of a solution containing 5 mM Tris-
 HCl (pH 7.5), 0.5 mM EDTA and 1 M NaCl, the non-biotinylated strand containing
 25 the radio-label was separated from the biotinylated strand by incubation in 50 µl of
 0.1 M NaOH for 10 min. The resultant supernatant was directly added to the
 hybridization solution described below.

[0013]

Incidentally, using the single-stranded (CAG)₅₅ probe prepared as described above, Southern blot analysis was carried out on the androgen receptor genes containing 9, 22, 43 and 51 CAG repeat units, respectively. As a result, the (CAG)₅₅ probe strongly hybridized with the genes having 43 and 51 CAG repeats units, respectively, but scarcely hybridized with the gene having 22 CAG repeat units, and did not hybridize at all with the gene having 9 CAG repeat units (K. Sanpei et al., Biochemical and Biophysical Research Communications, Vol.212, No.2, 1995, pp.341-346). Thus, by using this probe, hybridization may be selectively attained only with DNAs containing a number of (e.g., 35 or more) CAG repeat units if the hybridization conditions are appropriately selected.

[0014]

(2) Determination of Nucleotide Sequence of SCA2 Gene

Fig. 5 shows a pedigree chart of SCA2 patients. In this pedigree chart, males are represented by squares and females are represented by circles. SCA2 patients are represented by black squares or circles, and unaffected persons are represented by white squares or circles.

[0015]

High-molecular-weight genomic DNA (15 µg) was digested with 100 U of *TspEI* (Toyobo, Osaka, Japan), electrophoresed through 0.8% agarose gels and transferred to nitrocellulose membranes. The membranes were hybridized with the (CAG)₅₅ probe described above. Hybridization was performed in a solution containing 2.75 x SSPE (1 x SSPE=150 mM NaCl, 10 mM NaH₂PO₄, 1 mM EDTA), 50% formamide, 5 x Denhardt's solution, 100 ng/ml of sheared salmon sperm DNA and the (CAG)₅₅ probe (6 x 10⁶ cpm/ml) at 62°C for 18 hours. After the hybridization, the membranes were washed with 1 x SSC (150 mM NaCl, 15 mM sodium citrate) containing 0.5% SDS at 65°C for 0.5 hours. The membranes were

autoradiographed for 16 hours to Kodak Bio Max MS film at -70°C using an MS intensifying screen.

[0016]

As a result, 2.5 kbp *TspEI* fragment hybridized with the probe was detected
5 only in all of the SCA2 patients.

[0017]

Genomic DNA (270 µg) from an SCA2 patient (individual 7 in Fig. 5) was digested by *TspEI* and subjected to agarose gel electrophoresis. Genomic DNA fragments including the 2.5 kb *TspEI* fragment were cloned into an *EcoRI*-cleaved
10 λZAPII vector. The genomic library was screened using the (CAG)₅₅ probe under the hybridization condition described above. A genomic clone, *Tsp-1*, containing an expanded CAG repeat was isolated.

[0018]

After removal of the probe, the above-described genomic library was
15 screened again using the *Tsp-1* as a probe, which was labeled by the random priming. Hybridization was carried out in a solution containing 5 x SSC, 1 x Denhardt's solution, 10% dextran sulfate, 20 mM sodium phosphate, 400 µg/ml human placental DNA and the *Tsp-1* probe at 42°C for 18 hours. After the hybridization, the membranes were washed finally in 0.1 x SSC - 0.1% SDS at 52°C for 0.5 hours.
20 The membranes were autoradiographed for 24 hours to Kodak Bio Max MS films at -70°C using an MS intensifying screen. As a result, a genomic clone, *Tsp-2*, originated from a normal allele was isolated.

[0019]

The *SmaI*-*ApaI* fragment (630 bp) of *Tsp2* was sequenced and
25 oligonucleotides F-1 (5'-CCC TCA CCA TGT CGC TGA AGC-3') and R-1 (5'-CGA CGC TAG AAG GCC GCT G-3') were designed such that the CAG repeat units are sandwiched between the oligonucleotides (see Fig. 1). Using oligonucleotides F-1

and R-1 as probes, human procephalic cortex cDNA library (STRATAGENE) was screened. Hybridization was performed in a solution containing 6 x SSC, 10 x Denhardt's solution, 0.5% SDS, 0.05% sodium pyrophosphate, 100 ng/ml of sheared salmon sperm DNA and end-labeled oligonucleotide probes at 55°C for 18 hours.

5 After the hybridization, the membrane was finally washed with 6 x SSC containing 0.5% SDS and 0.05% sodium pyrophosphate at 55°C for 0.5 hours. A cDNA clone Fc1 with a size of 4.0 kb which hybridized with the both probes was obtained. The nucleotide sequences of Fc1, Tsp1 and Tsp2 were determined and compared. As a result, the nucleotide sequences in the vicinities of the CAG repeat units were
10 identical except for the number of the CAG repeat units. Restriction maps of Tsp1 and Tsp2, as well as the sizes and positions of Fc1 and other fragments hereinbelow described, are shown in Fig. 6. Using Fc1 or a fragment isolated by the screening later described as a probe, human cDNA libraries (human procephalic cortex, human fetal brain, human brain and brain stem) were screened to isolate cDNA clones Fc2, Fb14, B4, C6 and C19 (see Fig. 6). To identify the 5'-end of Fc1, 5'-RACE
15 (Frohman, M.A. et al, Proc. Natl. Acad. Sci. USA 85, 8998-9002 (1988)) was performed using 5'-RACE-Ready cDNA (Clontech, Palo Alto, CA, USA). Primer R-1 was used for the first PCR, and Primer R-2 (5'-CTT GCG GAC ATT GGC AGC C-3', see Fig. 1) was used for the second PCR. In both PCRs, F-1 (see Fig. 1) was
20 used as the forward primer. The 5'-RACE product (5R1) having the size of 350 bp was subcloned into pT7Blue T vector (pT7Blue T-vector (5R1)). The identification of 5R1 was confirmed by the overlapping with the nucleotide sequences of Fc1, Tsp1 and Tsp2. To identify the 3'-end of the cDNA, 3'-RACE was performed using 1 µg of poly(A)⁺mRNA extracted from human brain as a template and Primer F-13 (5'-
25 TTC TCT CAG CCA AAG CCT TCT ACT ACC-3', see Fig. 3) as a primer. The obtained 3'-RACE product (3R1) with a size of 1300 bp was subcloned into pT7Blue T vector (pT7Blue T-vector (3R1)).

[0020]

To investigate the 5'-end region of the cDNA, reverse transcription PCR (RT-PCR) was performed. That is, total RNAs extracted from an autopsy from human brain were digested by RNase-free DNase (PROMEGA) (Onodera, O. et al., Am. J. Hum. Geent. 57, 1050-1060(1995)). As the primers for the PCR, F1006 (5'-TAT CCG CAG CTC CGC TCC C-3', see Fig. 1) and R1002 (5'-AGC CGG GCC GAA ACG CGC CG-3') were used. PCR was performed in a solution with a total volume of 20 μ M, which contained 5 pmol each of the each primer, 10 mM Tris HCl (pH8.3), 50 mM KCl, 1.5 mM MgCl₂, 1.7M N,N,N-trimethylglycine, 200 μ M each of dATP, dCTP and TTP, 100 μ M of dGTP, 100 μ M of 7-deaza dGTP and 2.5 U of Taq polymerase (TAKARA SHUZO). After carrying out the initial denaturation at 96°C for 2 minutes, a cycle of a denaturation step at 96°C for 1 minute, an annealing step at 65°C for 1 minute and an extension step at 72°C for 1 minute were repeated 30 times, and a final extension step at 72°C for 5 minutes was performed, thereby carrying out the PCR. As a result, a clone 5R1 which extends upstream of 5R1 by 246 bp was obtained (see Fig. 6).

[0021]

In Fig. 6, the hollow regions in the Tsp1 and Tsp2 fragments indicate the regions which exist in SCA2 cDNAs. The hollow regions in the SCA2 cDNA shows coding regions. The CAG repeating regions are shown as solid boxes. Restriction sites TspE1 (T), NotI (N), Sac II (S), Sau3AI (Sa), Sma I (Sm), Eco52I (E52), Apa I (Ap), AccI (Ac), BamHI (B), XhoI (X), EcoRI (E) and Pst I (P) are shown. The size and position of each cDNA clone are shown below the consensus SCA2 cDNA.

[0022]

In this example, nucleotide sequences of double-stranded DNAs were determined by the dideoxynucleotide chain termination method (Sanger, F. et al.

Proc. Natl. Acad. Sci. USA 74, 5463-5467(1977); Chen E.Y. et al, DNA 4, 165-170 (1985)) using a double-stranded plasmid DNA as a template. To determine the nucleotide sequences of the CAG repeating regions and their flanking regions, genomic fragments containing the CAG repeating regions were amplified by PCR using biotinylated F-1 and RS-1 (5'-CCT CGG TGT CGC GGC GAC TTC C-3'). PCR was performed in a solution with a total volume of 25 μ l, which contained 0.25 μ M each of the each primer, 10 mM Tris HCl (pH8.3), 50 mM KCl, 2.0 mM $MgCl_2$, 1.7M N,N,N-trimethylglycine, 200 μ M each of dNTP, 200 ng of the genomic DNA and 1.25 U of Taq polymerase (TAKARA SHUZO). After carrying out initial denaturation at 95°C for 1 minute, a cycle of a denaturation step at 95°C for 2 minutes, an annealing step at 62°C for 1 minute and an extension step at 72°C for 1 minute was repeated 32 times, and a final extension step at 72°C for 5 minutes was performed, thereby carrying out the PCR. Biotinylated chains were recovered using streptavidin-coated magnetic beads and were directly sequenced.

[0023]

Based on the nucleotide sequences of the above-mentioned cDNA clones, a consensus SCA2 cDNA sequence with a length of 4351 bp excluding the poly A tail was determined (SEQ ID NO:1, Figs. 1-4, see Fig. 6). In SEQ ID NO: 1, the region from 4352nt to 4367nt is the poly A tail, and the number of "A" is not restricted to that shown in SEQ ID NO: 1. It was confirmed that the poly A tail exists at the same location in C19, B4 and 3R1 which were independent cDNA clones.

[0024]

Example 2 Measurement of CAG Repeat Units in Sample

Numbers of CAG repeat units were determined by polyacrylamide gel electrophoresis analysis of PCR products obtained using the primer pair of F-1 and R-1. PCR was performed in a total volume of 10 μ l containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.0 mM $MgCl_2$, 1.7 M N,N,N-trimethylglycine, 111KBq of $[\alpha$ -

^{32}P]dCTP (111 Tbq/mmol), 30 μM dCTP, and 200 μM each of dATP, dGTP and TTP, 0.25 μM each of the two primers, 200 ng of genomic DNA and 1.25 U of Taq DNA polymerase. After an initial 2-min denaturation at 95°C, PCR was performed for 32 cycles of 1-min denaturation at 95°C, 1-min annealing at 60°C and 1-min extension at 72°C, followed by a final extension at 72°C for 5 min. Sequence ladders obtained using the cloned genomic segments of the SCA2 gene, which contain various sizes of CAG repeats, were used as size markers. For normal alleles containing one or two CAA interruptions, the numbers of the CAA units were included in the CAG repeat size. For SCA2 alleles having expanded CAG region, the above-mentioned insert sequence immediately after the CAG region was not included in the size of the CAG region.

[0025]

By the above-described method, the numbers of the CAG repeat units of normal individuals (286 chromosomes) and 10 pedigrees of SCA2 patients (34 SCA2 chromosomes) were determined. The results are shown in Fig. 7. In Fig. 7, open bars indicate the results of the normal genes and solid bars indicate the results of the SCA2 genes.

[0026]

As is apparent from Fig. 7, in all of the normal genes, the numbers of the CAG repeat units were not more than 24, while in all of the SCA2 genes, they were not less than 35. Thus, it was confirmed that the cDNA identified as described above is the cDNA of the causative gene of SCA2.

[0027]

SEQUENCE LISTING

SEQ ID NO: 1

SEQUENCE LENGTH: 4367

SEQUENCE TYPE: nucleic acid

STRNDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION

TATCCGCACC TCCGCTCCCA CCCGGCGCCT CGGCGCGCCC GCCCTCCG ATG CGC TCA	57
Met Arg Ser	
1	
GCG GCC GCA GCT CCT CGG AGT CCC GCG GTG GCC ACC GAG TCT CGC CGC	105
Ala Ala Ala Ala Pro Arg Ser Pro Ala Val Ala Thr Glu Ser Arg Arg	
5 10 15	
TTC GCC GCA GCC AGG TGG CCC GGG TGG CGC TCG CTC CAG CGG CCG GCG	153
Phe Ala Ala Ala Arg Trp Pro Gly Trp Arg Ser Leu Gln Arg Pro Ala	
20 25 30 35	
CGG CGG AGC GGG CGG GGC GGC GGT GGC GCG GCC CCG GGA CCG TAT CCC	201
Arg Arg Ser Gly Arg Gly Gly Gly Gly Ala Ala Pro Gly Pro Tyr Pro	
40 45 50	
TCC GCC GCC CCT CCC CCG CCC GGC CCC GGC CCC CCT CCC TCC CGG CAG	249
Ser Ala Ala Pro Pro Pro Pro Gly Pro Gly Pro Pro Pro Ser Arg Gln	
55 60 65	
AGC TCG CCT CCC TCC GCC TCA GAC TGT TTT GGT AGC AAC GGC AAC GGC	297
Ser Ser Pro Pro Ser Ala Ser Asp Cys Phe Gly Ser Asn Gly Asn Gly	
70 75 80	
GGC GGC GCG TTT CGG CCC GGC TCC CGG CGG CTC CTT GGT CTC GGC GGC	345

1

230	235	240	
AGA GGT CGA AAC AGT AAC AAA GGA CTG CCT CAG TCT ACG ATT TCT TTT			825
Arg Gly Arg Asn Ser Asn Lys Gly Leu Pro Gln Ser Thr Ile Ser Phe			
245	250	255	
GAT GGA ATC TAT GCA AAT ATG AGG ATG GTT CAT ATA CTT ACA TCA GTT			873
Asp Gly Ile Tyr Ala Asn Met Arg Met Val His Ile Leu Thr Ser Val			
260	265	270	275
GTT GGC TCC AAA TGT GAA GTA CAA GTG AAA AAT GGA GGT ATA TAT GAA			921
Val Gly Ser Lys Cys Glu Val Gln Val Lys Asn Gly Gly Ile Tyr Glu			
280	285	290	
GGA GTT TTT AAA ACT TAC AGT CCG AAG TGT GAT TTG GTA CTT GAT GCC			969
Gly Val Phe Lys Thr Tyr Ser Pro Lys Cys Asp Leu Val Leu Asp Ala			
295	300	305	
GCA CAT GAG AAA AGT ACA GAA TCC AGT TCG GGG CCG AAA CGT GAA GAA			1017
Ala His Glu Lys Ser Thr Glu Ser Ser Ser Gly Pro Lys Arg Glu Glu			
310	315	320	
ATA ATG GAG AGT ATT TTG TTC AAA TGT TCA GAC TTT GTT GTG GTA CAG			1065
Ile Met Glu Ser Ile Leu Phe Lys Cys Ser Asp Phe Val Val Val Gln			
325	330	335	
TTT AAA GAT ATG GAC TCC AGT TAT GCA AAA AGA GAT GCT TTT ACT GAC			1113
Phe Lys Asp Met Asp Ser Ser Tyr Ala Lys Arg Asp Ala Phe Thr Asp			
340	345	350	355
TCT GCT ATC AGT GCT AAA GTG AAT GGC GAA CAC AAA GAG AAG GAC CTG			1161
Ser Ala Ile Ser Ala Lys Val Asn Gly Glu His Lys Glu Lys Asp Leu			
360	365	370	
GAG CCC TGG GAT GCA GGT GAA CTC ACA GCC AAT GAG GAA CTT GAG GCT			1209
Glu Pro Trp Asp Ala Gly Glu Leu Thr Ala Asn Glu Glu Leu Glu Ala			
375	380	385	

TTG GAA AAT GAC GTA TCT AAT GGA TGG GAT CCC AAT GAT ATG TTT CGA	1257
Leu Glu Asn Asp Val Ser Asn Gly Trp Asp Pro Asn Asp Met Phe Arg	
390 395 400	
TAT AAT GAA GAA AAT TAT GGT GTA GTG TCT ACG TAT GAT AGC AGT TTA	1305
Tyr Asn Glu Glu Asn Tyr Gly Val Val Ser Thr Tyr Asp Ser Ser Leu	
405 410 415	
TCT TCG TAT ACA GTG CCC TTA GAA AGA GAT AAC TCA GAA GAA TTT TTA	1353
Ser Ser Tyr Thr Val Pro Leu Glu Arg Asp Asn Ser Glu Glu Phe Leu	
420 425 430 435	
AAA CGG GAA GCA AGG GCA AAC CAG TTA GCA GAA GAA ATT GAG TCA AGT	1401
Lys Arg Glu Ala Arg Ala Asn Gln Leu Ala Glu Glu Ile Glu Ser Ser	
440 445 450	
GCC CAG TAC AAA GCT CGA GTG GCC CTG GAA AAC GAT GAT AGG AGT GAG	1449
Ala Gln Tyr Lys Ala Arg Val Ala Leu Glu Asn Asp Asp Arg Ser Glu	
455 460 465	
GAA GAA AAA TAC ACA GCA GTT CAG AGA AAT TCC AGT GAA CGT GAG GGG	1497
Glu Glu Lys Tyr Thr Ala Val Gln Arg Asn Ser Ser Glu Arg Glu Gly	
470 475 480	
CAC AGC ATA AAC ACT AGG GAA AAT AAA TAT ATT CCT CCT GGA CAA AGA	1545
His Ser Ile Asn Thr Arg Glu Asn Lys Tyr Ile Pro Pro Gly Gln Arg	
485 490 495	
AAT AGA GAA GTC ATA TCC TGG GGA AGT GGG AGA CAG AAT TCA CCG CGT	1593
Asn Arg Glu Val Ile Ser Trp Gly Ser Gly Arg Gln Asn Ser Pro Arg	
500 505 510 515	
ATG GGC CAG CCT GGA TCG GGC TCC ATG CCA TCA AGA TCC ACT TCT CAC	1641
Met Gly Gln Pro Gly Ser Gly Ser Met Pro Ser Arg Ser Thr Ser His	
520 525 530	
ACT TCA GAT TTC AAC CCG AAT TCT GGT TCA GAC CAA AGA GTA GTT AAT	1689

Thr Ser Asp Phe Asn Pro Asn Ser Gly Ser Asp Gln Arg Val Val Asn	
535 540 545	
GGA GGT GTT CCC TGG CCA TCG CCT TGC CCA TCT CCT TCC TCT CGC CCA	1737
Gly Gly Val Pro Trp Pro Ser Pro Cys Pro Ser Pro Ser Ser Arg Pro	
550 555 560	
CCT TCT CGC TAC CAG TCA GGT CCC AAC TCT CTT CCA CCT CGG GCA GCC	1785
Pro Ser Arg Tyr Gln Ser Gly Pro Asn Ser Leu Pro Pro Arg Ala Ala	
565 570 575	
ACC CCT ACA CGG CCG CCC TCC AGG CCC CCC TCG CGG CCA TCC AGA CCC	1833
Thr Pro Thr Arg Pro Pro Ser Arg Pro Pro Ser Arg Pro Ser Arg Pro	
580 585 590 595	
CCG TCT CAC CCC TCT GCT CAT GGT TCT CCA GCT CCT GTC TCT ACT ATG	1881
Pro Ser His Pro Ser Ala His Gly Ser Pro Ala Pro Val Ser Thr Met	
600 605 610	
CCT AAA CGC ATG TCT TCA GAA GGG CCT CCA AGG ATG TCC CCA AAG GCC	1929
Pro Lys Arg Met Ser Ser Glu Gly Pro Pro Arg Met Ser Pro Lys Ala	
615 620 625	
CAG CGA CAT CCT CGA AAT CAC AGA GTT TCT GCT GGG AGG GGT TCC ATA	1977
Gln Arg His Pro Arg Asn His Arg Val Ser Ala Gly Arg Gly Ser Ile	
630 635 640	
TCC AGT GGC CTA GAA TTT GTA TCC CAC AAC CCA CCC AGT GAA GCA GCT	2025
Ser Ser Gly Leu Glu Phe Val Ser His Asn Pro Pro Ser Glu Ala Ala	
645 650 655	
ACT CCT CCA GTA GCA AGG ACC AGT CCC TCG GGG GGA ACG TGG TCA TCA	2073
Thr Pro Pro Val Ala Arg Thr Ser Pro Ser Gly Gly Thr Trp Ser Ser	
660 665 670 675	
GTG GTC AGT GGG GTT CCA AGA TTA TCC CCT AAA ACT CAT AGA CCC AGG	2121
Val Val Ser Gly Val Pro Arg Leu Ser Pro Lys Thr His Arg Pro Arg	

680	685	690	
TCT CCC AGA CAG AAC AGT ATT GGA AAT ACC CCC AGT GGG CCA GTT CTT	2169		
Ser Pro Arg Gln Asn Ser Ile Gly Asn Thr Pro Ser Gly Pro Val Leu			
695	700	705	
GCT TCT CCC CAA GCT GGT ATT ATT CCA ACT GAA GCT GTT GCC ATG CCT	2217		
Ala Ser Pro Gln Ala Gly Ile Ile Pro Thr Glu Ala Val Ala Met Pro			
710	715	720	
ATT CCA GCT GCA TCT CCT ACG CCT GCT AGT CCT GCA TCG AAC AGA GCT	2265		
Ile Pro Ala Ala Ser Pro Thr Pro Ala Ser Pro Ala Ser Asn Arg Ala			
725	730	735	
GTT ACC CCT TCT AGT GAG GCT AAA GAT TCC AGG CTT CAA GAT CAG AGG	2313		
Val Thr Pro Ser Ser Glu Ala Lys Asp Ser Arg Leu Gln Asp Gln Arg			
740	745	750	755
CAG AAC TCT CCT GCA GGG AAT AAA GAA AAT ATT AAA CCC AAT GAA ACA	2361		
Gln Asn Ser Pro Ala Gly Asn Lys Glu Asn Ile Lys Pro Asn Glu Thr			
760	765	770	
TCA CCT AGC TTC TCA AAA GCT GAA AAC AAA GGT ATA TCA CCA GTT GTT	2409		
Ser Pro Ser Phe Ser Lys Ala Glu Asn Lys Gly Ile Ser Pro Val Val			
775	780	785	
TCT GAA CAT AGA AAA CAG ATT GAT GAT TTA AAG AAA TTT AAG AAT GAT	2457		
Ser Glu His Arg Lys Gln Ile Asp Asp Leu Lys Lys Phe Lys Asn Asp			
790	795	800	
TTT AGG TTA CAG CCA AGT TCT ACT TCT GAA TCT ATG GAT CAA CTA CTA	2505		
Phe Arg Leu Gln Pro Ser Ser Thr Ser Glu Ser Met Asp Gln Leu Leu			
805	810	815	
AAC AAA AAT AGA GAG GGA GAA AAA TCA AGA GAT TTG ATC AAA GAC AAA	2553		
Asn Lys Asn Arg Glu Gly Glu Lys Ser Arg Asp Leu Ile Lys Asp Lys			
820	825	830	835

ATT GAA CCA AGT GCT AAG GAT TCT TTC ATT GAA AAT AGC AGC AGC AAC	2601
Ile Glu Pro Ser Ala Lys Asp Ser Phe Ile Glu Asn Ser Ser Ser Asn	
840 845 850	
TGT ACC AGT GGC AGC AGC AAG CCG AAT AGC CCC AGC ATT TCC CCT TCA	2649
Cys Thr Ser Gly Ser Ser Lys Pro Asn Ser Pro Ser Ile Ser Pro Ser	
855 860 865	
ATA CTT AGT AAC ACG GAG CAC AAG AGG GGA CCT GAG GTC ACT TCC CAA	2697
Ile Leu Ser Asn Thr Glu His Lys Arg Gly Pro Glu Val Thr Ser Gln	
870 875 880	
GGG GTT CAG ACT TCC AGC CCA GCA TGT AAA CAA GAG AAA GAC GAT AAG	2745
Gly Val Gln Thr Ser Ser Pro Ala Cys Lys Gln Glu Lys Asp Asp Lys	
885 890 895	
GAA GAG AAG AAA GAC GCA GCT GAG CAA GTT AGG AAA TCA ACA TTG AAT	2793
Glu Glu Lys Lys Asp Ala Ala Glu Gln Val Arg Lys Ser Thr Leu Asn	
900 905 910 915	
CCC AAT GCA AAG GAG TTC AAC CCA CGT TCC TTC TCT CAG CCA AAG CCT	2841
Pro Asn Ala Lys Glu Phe Asn Pro Arg Ser Phe Ser Gln Pro Lys Pro	
920 925 930	
TCT ACT ACC CCA ACT TCA CCT CGG CCT CAA GCA CAA CCT AGC CCA TCT	2889
Ser Thr Thr Pro Thr Ser Pro Arg Pro Gln Ala Gln Pro Ser Pro Ser	
935 940 945	
ATG GTG GGT CAT CAA CAG CCA ACT CCA GTT TAT ACT CAG CCT GTT TGT	2937
Met Val Gly His Gln Gln Pro Thr Pro Val Tyr Thr Gln Pro Val Cys	
950 955 960	
TTT GCA CCA AAT ATG ATG TAT CCA GTC CCA GTG AGC CCA GGC GTG CAA	2985
Phe Ala Pro Asn Met Met Tyr Pro Val Pro Val Ser Pro Gly Val Gln	
965 970 975	
CCT TTA TAC CCA ATA CCT ATG ACG CCC ATG CCA GTG AAT CAA GCC AAG	3033

—

1125	1130	1135	
CTG CAC CCA CAT ACT CCA CAC CCT CAG CCT TCA GCT ACC CCC ACT GGA			3513
Leu His Pro His Thr Pro His Pro Gln Pro Ser Ala Thr Pro Thr Gly			
1140	1145	1150	1155
CAG CAG CAA AGC CAA CAT GGT GGA AGT CAT CCT GCA CCC AGT CCT GTT			3561
Gln Gln Gln Ser Gln His Gly Gly Ser His Pro Ala Pro Ser Pro Val			
	1160	1165	1170
CAG CAC CAT CAG CAC CAG GCC GCC CAG GCT CTC CAT CTG GCC AGT CCA			3609
Gln His His Gln His Gln Ala Ala Gln Ala Leu His Leu Ala Ser Pro			
	1175	1180	1185
CAG CAG CAG TCA GCC ATT TAC CAC GCG GGG CTT GCG CCA ACT CCA CCC			3657
Gln Gln Gln Ser Ala Ile Tyr His Ala Gly Leu Ala Pro Thr Pro Pro			
	1190	1195	1200
TCC ATG ACA CCT GCC TCC AAC ACG CAG TCG CCA CAG AAT AGT TTC CCA			3705
Ser Met Thr Pro Ala Ser Asn Thr Gln Ser Pro Gln Asn Ser Phe Pro			
	1205	1210	1215
GCA GCA CAA CAG ACT GTC TTT ACG ATC CAT CCT TCT CAC GTT CAG CCG			3753
Ala Ala Gln Gln Thr Val Phe Thr Ile His Pro Ser His Val Gln Pro			
1220	1225	1230	1235
GCG TAT ACC AAC CCA CCC CAC ATG GCC CAC GTA CCT CAG GCT CAT GTA			3801
Ala Tyr Thr Asn Pro Pro His Met Ala His Val Pro Gln Ala His Val			
	1240	1245	1250
CAG TCA GGA ATG GTT CCT TCT CAT CCA ACT GCC CAT GCG CCA ATG ATG			3849
Gln Ser Gly Met Val Pro Ser His Pro Thr Ala His Ala Pro Met Met			
	1255	1260	1265
CTA ATG ACG ACA CAG CCA CCC GGC GGT CCC CAG GCC GCC CTC GCT CAA			3897
Leu Met Thr Thr Gln Pro Pro Gly Gly Pro Gln Ala Ala Leu Ala Gln			
	1270	1275	1280

AGT GCA CTA CAG CCC ATT CCA GTC TCG ACA ACA GCG CAT TTC CCC TAT 3945
 Ser Ala Leu Gln Pro Ile Pro Val Ser Thr Thr Ala His Phe Pro Tyr
 1285 1290 1295
 ATG ACG CAC CCT TCA GTA CAA GCC CAC CAC CAA CAG CAG TTG 3987
 Met Thr His Pro Ser Val Gln Ala His His Gln Gln Gln Leu
 1300 1305 1310
 TAAGGCTGCC CTGGAGGAAC CGAAAGGCCA AATTCCTCC TCCCTTCTAC TGCTTCTACC 4047
 AACTGGAAGC ACAGAAAAC TGAATTTTCAT TTATTTTGT TTTAAAATAT ATATGTTGAT 4107
 TTCTTGTAAC ATCCAATAGG AATGCTAACA GTTCACTTGC AGTGAAGAT ACTTGGACCG 4167
 AGTAGAGGCA TTTAGGAACT TGGGGGCTAT TCCATAATTC CATATGCTGT TTCAGAGTCC 4227
 CGCAGGTACC CCAGCTCTGC TTGCCGAAAC TGAAGTTAT TTATTTTTTA ATAACCCTTG 4287
 AAAGTCATGA ACACATCAGC TAGCAAAAGA AGTAACAAGA GTGATTCTTG CTGCTATTAC 4347
 TGCTAAAAAA AAAAAAAAAA 4367

[BRIEF DESCRIPTION OF THE DRAWINGS]

[Fig. 1]

A drawing which shows the nucleotide sequence of the cDNA fragment according to the present invention together with the amino acid sequence encoded thereby, which nucleotide sequence was determined in the Examples of the present invention.

[Fig. 2]

A drawing which shows the continuation of Fig. 1.

[Fig. 3]

A drawing which shows the continuation of Fig. 2.

[Fig. 4]

A drawing which shows the continuation of Fig. 3.

[Fig. 5]

A pedigree chart of the SCA2 patients who donated the genomic DNAs used in

the Examples.

[Fig. 6]

A drawing which shows the sizes, positions and restriction sites of the genomic DNA fragments Tsp1 and Tsp2, and SCA2 cDNA obtained in the Examples of the present invention, as well as the size and position of each of the obtained cDNA fragment.

[Fig. 7]

A drawing which shows the distribution of the numbers of the CAG repeat units in the normal and SCA2 genes, which were measured by using the (CAG)₅₅ probe.

1 TATCCGCACCTCCGCTCCACCCGGCGCTCGGGCGCGCCCGCCCTCCGATGCGCTCAGCG
 1 F-1006 M R S A
 61 GCCGCAGCTCCTCGGAGTCCCGCGGTGGCCACCGAGTCTCGCCGCTTCGCCGCAGCCAGG
 5 A A A P R S P A V A T E S R R F A A A R
 121 TGGCCCCGGTGGCGCTCGCTCCAGCGGCCGGCGCGGAGCGGGCGGGCGGGCGGTGGC
 25 W P G W R S L Q R P A R R S G R G G G G
 181 GCGGCCCCGGGACCGTATCCCTCCGCGCCCTCCCCCGCCCGGCCCCGGCCCCCTCCC
 45 A A P G P Y P S A A P P P P G P G P P P
 241 TCCCGGCAGAGCTCGCCTCCCTCCGCTCAGACTGTTTTGGTAGCAACGGCAACGGCGGC
 65 S R Q S S P P S A S D C F G S N G N G G
 301 GCGCGTTTCGGCCCCGGCTCCCGCGGGCTCCTTGGTCTCGGCGGGCCTCCCCGCCCTTC
 85 G A F R P G S R R L L G L G G P P R P F
 361 GTCGTCGTCCTTCTCCCCCTCGCCAGCCCGGGCGCCCTCCGGCCGCGCCAACCGCGGCC
 105 V V V L L P L A S P G A P P A A P T R A
 421 TCCCGCTCGGCGCCCGTGCCTCCCGCCCGCTTCCGGCGTCTCCTTGGCGCGCCCGGCT
 125 S P L G A R A S P P R S G V S L A R P A
 481 CCCGGCTGTCCCGCCCGGCGTGCAGCCGGTGTATGGGCCCCCTACCATGTCGTGAAG
 145 P G C P R P A C E P V Y G F-1 P L T M S L K
 541 CCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAG
 165 P Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q
 601 CAGCAGCAGCAGCCCGCCCGCGGCTGCCAATGTCCGCAAGCCCGGGCGGCGGCCTT
 185 Q Q Q Q P P P A A A N V R K R-2 P G S G L
 661 CTAGCGTCGCCCCCGCGCGCCTTCGCCGTCTCGTCTCGGTCTCCTCGTCTCGGCC
 205 L A S R-1 A A A P S P S S S S V S S S S A
 721 ACGGCTCCCTCCTCGGTGGTTCGCGCGACCTCCGGCGCGGGAGGCCCGGCTGGGCAGA
 225 T A P S S V V A A T S G G G R P G L G R
 781 GGTCGAAACAGTAACAAAGGACTGCCTCAGTCTACGATTTCTTTTGATGGAATCTATGCA
 245 G R N S N K G L P Q S T I S F D G I Y A
 841 AATATGAGGATGGTTTCATATACTTACATCAGTTGTTGGCTCCAAATGTGAAGTACAAGTG
 265 N M R M V H I L T S V V G S K C E V Q V
 901 AAAAATGGAGGTATATATGAAGGAGTTTTTAAACTTACAGTCCGAAGTGTGATTTGGTA
 285 K N G G I Y E G V F K T Y S P K C D L V
 961 CTTGATGCCGCACATGAGAAAAGTACAGAATCCAGTTCCGGGGCCGAAACGTGAAGAAATA
 305 L D A A H E K S T E S S S G P K R E E I
 1021 ATGGAGAGTATTTTGTTCAAATGTTTCAGACTTTGTTGTGCTACAGTTTAAAGATATGGAC
 325 M E S I L F K C S D F V V V Q F K D M D
 1081 TCCAGTTATGCAAAAAGAGATGCTTTTACTGACTCTGCTATCAGTGCTAAAGTGAATGGC
 345 S S Y A K R D A F T D S A I S A K V N G
 1141 GAACACAAAGAGAAGGACCTGGAGCCCTGGGATGCAGGTGAACTCACAGCCAATGAGGAA
 365 E H K E K D L E P W D A G E L T A N E E
 1201 CTTGAGGCTTTGGAAAATGACGTATCTAATGGATGGGATCCCAATGATATGTTTCGATAT
 385 L E A L E N D V S N G W D P N D M F R Y
 1261 AATGAAGAAAATTATGGTGTAGTGTCTACGTATGATAGCAGTTTATCTTCGTATACAGTG
 405 N E E N Y G V V S T Y D S S L S S Y T V

Fig. 1

1321 CCCTTAGAAAGAGATAACTCAGAAGAATTTTAAACGGGAAGCAAGGGCAAACCAGTTA
 425 P L E R D N S E E F L K R E A R A N Q L
 1381 GCAGAAGAAATTGAGTCAAGTGCCCAAGTACAAAGCTCGAGTGGCCCTGGAAAACGATGAT
 445 A E E I E S S A Q Y K A R V A L E N D D
 1441 AGGAGTGAGGAAGAAAAATACACAGCAGTTCAGAGAAATTCCAGTGAACGTGAGGGGCAC
 465 R S E E E K Y T A V Q R N S S E R E G H
 1501 AGCATAAACACTAGGGAAAATAAATATATTCCTCCTGGACAAAGAAATAGAGAAGTCATA
 485 S I N T R E N K Y I P P G Q R N R E V I
 1561 TCCTGGGGAAAGTGGGAGACAGAATTCACCGCGTATGGGCCAGCCTGGATCGGGCTCCATG
 505 S W G S G R Q N S P R M G Q P G S G S M
 1621 CCATCAAGATCCACTTCTCACACTTCAGATTTCAACCCGAATTCTGGTTTCAGACCAAAGA
 525 P S R S T S H T S D F N P N S G S D Q R
 1681 GTAGTTAATGGAGGTGTTCCCTGGCCATCGCCTTGCCCATCTCCTTCTCGCCACCT
 545 V V N G G V P W P S P C P S P S S R P P
 1741 TCTCGCTACCAGTCAGGTCCCAACTCTCTCCACCTCGGGCAGCCACCCCTACACGGCCG
 565 S R Y Q S G P N S L P P R A A T P T R P
 1801 CCCTCCAGGCCCCCTCGCGGCCATCCAGACCCCGTCTCACCCCTCTGCTCATGGTTCT
 585 P S R P P S R P S R P P S H P S A H G S
 1861 CCAGCTCCTGTCTCTACTATGCCTAAACGCATGTCTTCAGAAGGGCCTCCAAGGATGTCC
 605 P A P V S T M P K R M S S E G P P R M S
 1921 CCAAAGGCCCAGCGACATCCTCGAAATCACAGAGTTTCTGCTGGGAGGGGTTCCATATCC
 625 P K A Q R H P R N H R V S A G R G S I S
 1981 AGTGGCCTAGAATTTGTATCCCACAACCCACCCAGTGAAGCAGCTACTCCTCCAGTAGCA
 645 S G L E F V S H N P P S E A A T P P V A
 2041 AGGACCAGTCCCTCGGGGGGAACGTGGTCATCAGTGGTCAGTGGGGTTCCAAGATTATCC
 665 R T S P S G G T W S S V V S G V P R L S
 2101 CCTAAACTCATAGACCCAGGTCTCCAGACAGAACAGTATTGGAAATACCCCCAGTGGG
 685 P K T H R P R S P R Q N S I G N T P S G
 2161 CCAGTTCTTGCTTCTCCCCAAGCTGGTATTATTCCAAGTGAAGCTGTTGCCATGCCTATT
 705 P V L A S P Q A G I I P T E A V A M P I
 2221 CCAGCTGCATCTCCTACGCCTGCTAGTCTGCATCGAACAGAGCTGTTACCCCTTCTAGT
 725 P A A S P T P A S P A S N R A V T P S S
 2281 GAGGCTAAAGATTCCAGGCTTCAAGATCAGAGGCAGAACTCTCCTGCAGGGAATAAGAA
 745 E A K D S R L Q D Q R Q N S P A G N K E
 2341 AATATTAAACCAATGAAACATCACCTAGCTTCTCAAAGCTGAAAACAAAGGTATATCA
 765 N I K P N E T S P S F S K A E N K G I S
 2401 CCAGTTGTTTCTGAACATAGAAAACAGATTGATGATTTAAAGAAATTTAAGAATGATTTT
 785 P V V S E H R K Q I D D L K K F K N D F
 2461 AGGTTACAGCCAAGTTCTACTTCTGAATCTATGGATCAACTACTAAACAAAATAGAGAG
 805 R L Q P S S T S E S M D Q L L N K N R E
 2521 GGAGAAAAATCAAGAGATTTGATCAAAGACAAAATGAACCAAGTGCTAAGGATTCTTTT
 825 G E K S R D L I K D K I E P S A K D S F
 2581 ATTGAAAATAGCAGCAGCAACTGTACCAAGTGGCAGCAGCAAGCCGAATAGCCCCAGCATT
 845 I E N S S S N C T S G S S K P N S P S I

Fig. 2

2641 TCCCCTTCAATACTTAGTAACACGGAGCACAAGAGGGGACCTGAGGTCACTTCCCAAGGG
 865 S P S I L S N T E H K R G P E V T S Q G
 2701 GTTCAGACTTCCAGCCCAGCATGTAAACAAGAGAAAGACGATAAGGAAGAGAAGAAAGAC
 885 V Q T S S P A C K Q E K D D K E E K K D
 2761 GCAGCTGAGCAAGTTAGGAAATCAACATTGAATCCCAATGCAAAGGAGTTCAACCCACGT
 905 A A E Q V R K S T L N P N A K E F N P R
 2821 TCCTTCTCTCAGCCAAAGCCTTCTACTACCCCAACTTCACCTCGGCCTCAAGCACAACCT
 F-13
 925 S F S Q P K P S T T P T S P R P Q A Q P
 2881 AGCCCATCTATGGTGGGTCAACAGCCAACTCCAGTTTATACTCAGCCTGTTGTTTT
 945 S P S M V G H Q Q P T P V Y T Q P V C F
 2941 GCACCAAATATGATGTATCCAGTCCCAGTGAGCCCAGGCGTGCAACCTTTATACCCAATA
 965 A P N M M Y P V P V S P G V Q P L Y P I
 3001 CCTATGACGCCCATGCCAGTGAATCAAGCCAAGACATATAGAGCAGTACCAAATATGCCC
 985 P M T P M P V N Q A K T Y R A V P N M P
 3061 CAACAGCGGCAAGACCAGCATCATCAGAGTGCCATGATGCACCCAGCGTCAGCAGCGGGC
 1005 Q Q R Q D Q H H Q S A M M H P A S A A G
 3121 CCACCGATTGCAGCCACCCACAGCTTACTCCACGCAATATGTTGCCTACAGTCCTCAG
 1025 P P I A A T P P A Y S T Q Y V A Y S P Q
 3181 CAGTTCCAAATCAGCCCCTTGTTTCAGCATGTGCCACATTATCAGTCTCAGCATCCTCAT
 1045 Q F P N Q P L V Q H V P H Y Q S Q H P H
 3241 GTCTATAGTCCTGTAATACAGGTAATGCTAGAATGATGGCACCACCAACACAGCCCCAG
 1065 V Y S P V I Q G N A R M M A P P T H A Q
 3301 CCTGGTTTAGTATCTTCTTCAGCAACTCAGTACGGGGCTCATGAGCAGACGCATGCGATG
 1085 P G L V S S S A T Q Y G A H E Q T H A M
 3361 TATGCATGTCCCAAATTACCATAACAAGGAGACAAGCCCTTCTTCTACTTTGCCATT
 1105 Y A C P K L P Y N K E T S P S F Y F A I
 3421 TCCACGGGCTCCCTTGCTCAGCAGTATGCGCACCCCTAACGCTACCCTGCACCCACATACT
 1125 S T G S L A Q Q Y A H P N A T L H P H T
 3481 CCACACCCTCAGCCTTCAGCTACCCCACTGGACAGCAGCAAAGCCAACATGGTGGAAGT
 1145 P H P Q P S A T P T G Q Q Q S Q H G G S
 3541 CATCCTGCACCCAGTCCTGTTTCAGCACCATCAGCACCCAGGCCGCCAGGCTCTCCATCTG
 1165 H P A P S P V Q H H Q H Q A A Q A L H L
 3601 GCCAGTCCACAGCAGCAGTCAGCCATTTACCACGCGGGGCTTGCGCCAACCTCCACCCCTCC
 1185 A S P Q Q Q S A I Y H A G L A P T P P S
 3661 ATGACACCTGCCTCCAACACGCGAGTCGCCACAGAATAGTTTCCCAGCAGCACAACAGACT
 1205 M T P A S N T Q S P Q N S F P A A Q Q T
 3721 GTCTTTACGATCCATCCTTCTCAGTTTCAGCCGGCGTATACCAACCCACCCACATGGCC
 1225 V F T I H P S H V Q P A Y T N P P H M A
 3781 CACGTACCTCAGGCTCATGTACAGTCAGGAATGGTTCTTCTCATCCAACCTGCCCATGCG
 1245 H V P Q A H V Q S G M V P S H P T A H A
 3841 CCAATGATGCTAATGACGACACAGCCACCCGGCGGTCCCCAGGCCGCCCTCGCTCAAAGT
 1265 P M M L M T T Q P P G G P Q A A L A Q S
 3901 GCACTACAGCCCATTCCAGTCTCGACAACAGCGCATTTCCCCTATATGACGCACCCCTTCA
 1285 A L Q P I P V S T T A H F P Y M T H P S
 3961 GTACAAGCCCACCACCAACAGCAGTTGTAAGGCTGCCCTGGAGGAACCGAAAGGCCAAAT
 1305 V Q A H H Q Q Q L *

Fig. 3

4021 TCCCTCCTCCCTTCTACTGCTTCTACCAACTGGAAGCACAGAAAAGTAGAATTTTCATTTA
 4081 TTTTGTTTTTAAATATATATATGTTGATTTCTTGTAACATCCAATAGGAATGCTAACAGTT
 4141 CACTTGCAAGTGAAGATACTTGGACCGAGTAGAGGCATTTAGGAAGTTGGGGGCTATTCC
 4201 ATAATTCCATATGCTGTTTCAGAGTCCCGCAGGTACCCAGCTCTGCTTGCCGAAACTGG
 4261 AAGTTATTTATTTTTTAATAACCCTTGAAAGTCATGAACACATCAGCTAGCAAAAGAAGT
 4321 AACAAAGAGTGATTCTTGCTGCTATTACTGCT (A)_n

Fig. 4

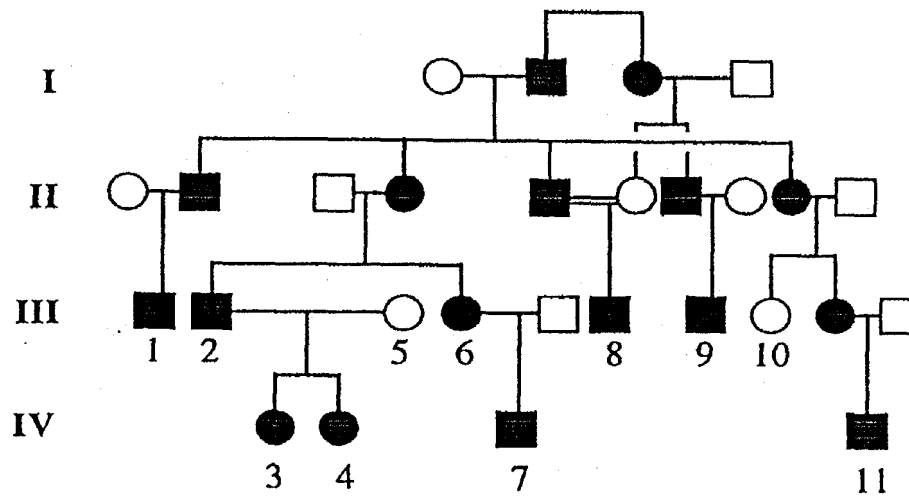


Fig. 5



Fig. 6

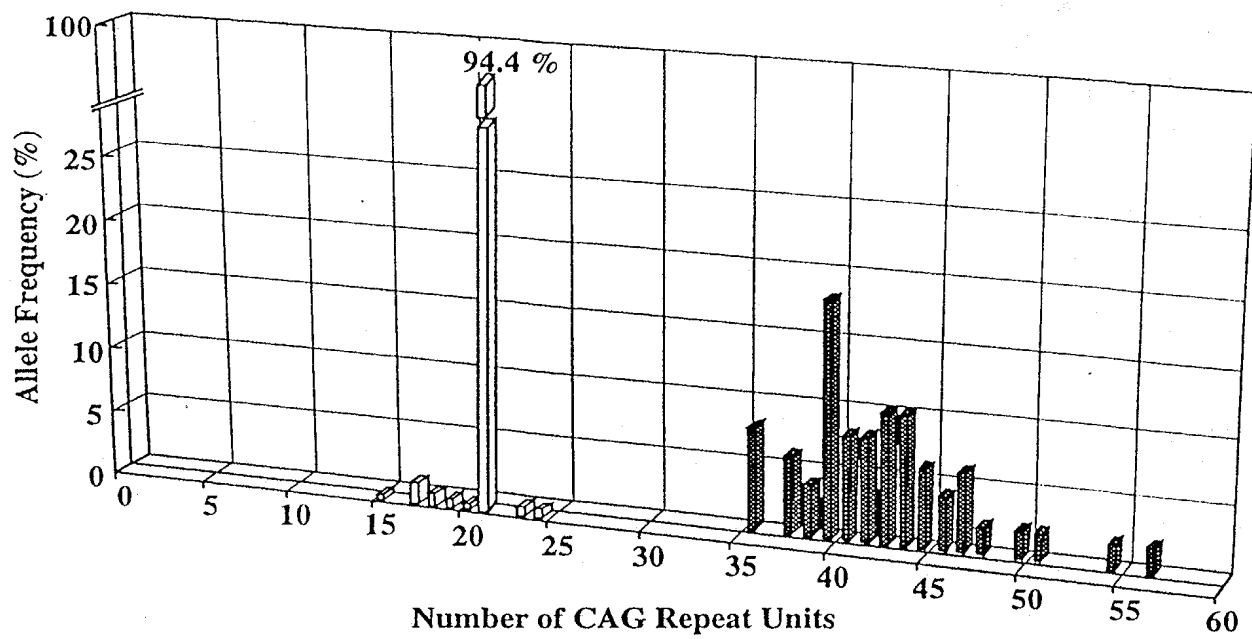


Fig. 7